Molecular Plant Pathology

MOLECULAR PLANT PATHOLOGY (2016) 17(6), 860–874 DOI: 10.1111/mpp.12334

Dual RNA-seq reveals *Meloidogyne graminicola* transcriptome and candidate effectors during the interaction with rice plants

ANNE-SOPHIE PETITOT[1,*], ALEXIS DEREEPER[1], MAWUSSE AGBESSI[1], CORINNE DA SILVA[2], JULIE GUY[2], MORGANE ARDISSON[3] AND DIANA FERNANDEZ[1]

[1]*IRD, UMR IRD-Cirad-UM2 Interactions Plantes-Microbes-Environnement, 34394, Montpellier Cedex 5, France*
[2]*CEA, Institut de Génomique, GENOSCOPE - Centre National de Séquençage, 91057, Evry Cedex, France*
[3]*INRA, UMR Amélioration Génétique et Adaptation des Plantes, 34060, Montpellier Cedex 1, France*

SUMMARY

Root-knot nematodes secrete proteinaceous effectors into plant tissues to facilitate infection by suppressing host defences and reprogramming the host metabolism to their benefit. *Meloidogyne graminicola* is a major pest of rice (*Oryza sativa*) in Asia and Latin America, causing important crop losses. The goal of this study was to identify *M. graminicola* pathogenicity genes expressed during the plant–nematode interaction. Using the dual RNA-sequencing (RNA-seq) strategy, we generated transcriptomic data of *M. graminicola* samples covering the pre-parasitic J2 stage and five parasitic stages in rice plants, from the parasitic J2 to the adult female. In the absence of a reference genome, a *de novo M. graminicola* transcriptome of 66 396 contigs was obtained from those reads that were not mapped on the rice genome. Gene expression profiling across the *M. graminicola* life cycle revealed key genes involved in nematode development and provided insights into the genes putatively associated with parasitism. The development of a 'secreted protein prediction' pipeline revealed a typical set of proteins secreted by nematodes, as well as a large number of cysteine-rich proteins and putative nuclear proteins. Combined with expression data, this pipeline enabled the identification of 15 putative effector genes, including two homologues of well-characterized effectors from cyst nematodes (CLE-like and VAP1) and a metallothionein. The localization of gene expression was assessed by *in situ* hybridization for a subset of candidates. All of these data represent important molecular resources for the elucidation of *M. graminicola* biology and for the selection of potential targets for the development of novel control strategies for this nematode species.

Keywords: dual RNA-seq, effector genes, expression profiling, *Meloidogyne graminicola*, *Oryza sativa*, plant-parasitic nematode.

* *Correspondence*: Email: anne-sophie.petitot@ird.fr

INTRODUCTION

Root-knot nematodes (RKNs) (*Meloidogyne* species) are microscopic worms that infect plant root systems and cause large economic losses in cultivated plants. They are obligate sedentary endoparasites which cause symptoms of swelling or galls on plant root systems. In order to complete their life cycle, RKNs must invade the roots of a susceptible plant, where they induce the differentiation of host cells into hypertrophied, multinucleate and metabolically active feeding cells, named giant cells. This feeding structure serves as a constant food source to the nematode, allowing its development into a reproductive female. Females lay egg-masses, containing hundreds to thousands of eggs, directly into the rhizosphere or within the gall tissue. After hatching, new juveniles arising from the eggs start a new reproduction cycle in another root from the same host, or from a new one. Cyst nematodes (*Globodera* and *Heterodera* species) are other sedentary plant-parasitic nematodes that cause significant damage to cultivated plants. They mainly differ from RKNs by their feeding site, which is composed of a large multinucleate syncytium originating from the incorporation of host cells surrounding the initial syncytial cell targeted by the nematode.

It has long been recognized that plant-parasitic nematodes, like other pathogens, successfully infect their hosts through the release of virulence effectors, primarily proteins, to facilitate infection by reprogramming the host metabolism and preventing the execution of plant defence responses. Effectors with diverse functions are produced to facilitate the nematode life cycle inside the plant (for a review, see Haegeman *et al.*, 2012; Mitchum *et al.*, 2013). Transcriptomic and proteomic approaches have allowed the identification of putative secreted proteins from various *Meloidogyne incognita* developmental stages (Bellafiore *et al.*, 2008; Huang *et al.*, 2003). More recently, the use of high-throughput sequencing technologies has expanded the identification of the effectors secreted by sedentary nematodes. Genome data mining of *M. incognita* and *Meloidogyne hapla*, the two *Meloidogyne* genomes sequenced so far (Abad *et al.*, 2008; Opperman *et al.*, 2008), combined with RNA-silencing experiments, has enabled the validation of 12 novel effector genes involved in *M. incognita* parasitism (Danchin *et al.*, 2013). In another study performed on

© 2015 BSPP AND JOHN WILEY & SONS LTD

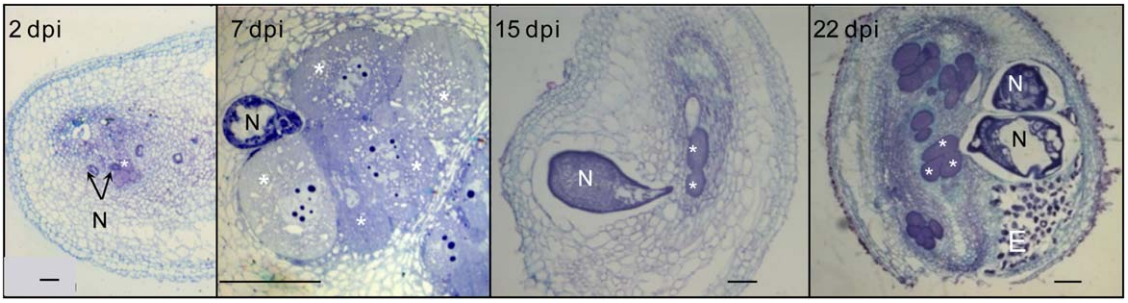

Fig. 1 *Meloidogyne graminicola* life cycle in rice roots (*Oryza sativa* cv. Nipponbare). Root cross-sections (10 μm) obtained at 2, 7, 15 and 22 days after nematode infection (days post-infection, dpi) were stained with toluidine blue (N, nematode; asterisk, giant cells; E, eggs).

M. incognita, Rutter *et al.* (2014a) sequenced specific mRNAs isolated from gland cells dissected at several parasitic stages and identified 18 putative effectors by a combination of *in silico* and molecular approaches. Similar strategies have also been employed to discover novel effectors from the recently sequenced genome of the cyst nematode *Globodera pallida*. Sequence annotation combined with RNA-sequencing (RNA-seq) data obtained at eight *G. pallida* life stages allowed the prediction of novel candidate effectors and the description of key genes involved in root invasion and the establishment of feeding sites (Cotton *et al.*, 2014; Thorpe *et al.*, 2014).

The RKN *Meloidogyne graminicola* is a major pest of rice (*Oryza sativa* L.) in Asia and Central America (Kyndt *et al.*, 2014). Severe infections result in a stunted root system, preventing the development of the rice plant. Important yield reductions have been reported, up to 70% in an infected field (Bridge *et al.*, 2005). The nematode has a life cycle of around 14–21 days depending on the temperature and flooding conditions (Kyndt *et al.*, 2014). Histological analysis of *M. graminicola* development in rice (cv. Nipponbare) roots showed the formation of giant cells from 2 days post-infection (dpi), swollen parasitic juveniles (stage J2) feeding on giant cells from 4 dpi, nematodes at stages J3 and J4 from 12 dpi, reaching the female stage from 15 dpi, and egg-laying females from 18 to 22 dpi (Nguyen *et al.*, 2014; Fig. 1). One feature of *M. graminicola*, compared with other *Meloidogyne* species, is that female bodies and egg-masses stay embedded within the gall tissues. Thus, its life cycle can be completed without leaving the host, and this is assumed to be an adaptation to flooded conditions (Bridge *et al.*, 2005).

No *M. graminicola* reference genome is available to date. A first 454-transcriptome of 10 257 contigs was established from pre-parasitic J2 larvae, and RNA-seq performed on rice-infected tissues at 3 and 7 dpi allowed around 6000 additional *M. graminicola* contigs to be obtained (Haegeman *et al.*, 2013). A series of putative effector genes expressed in the oesophageal glands of the nematode were identified, together with four genes with a lethal RNA interference (RNAi) phenotype. Only very few homologues to effectors identified in related *Meloidogyne* species were found, suggesting that this *M. graminicola* transcriptome was incomplete or that *M. graminicola* may use a divergent set of effectors (Haegeman *et al.*, 2013). Deciphering *M. graminicola* gene expression patterns throughout its life cycle should provide better insights into nematode pathogenicity mechanisms and help to identify new candidate effectors. To obtain an extensive set of genes involved in parasitism, we generated large-scale transcriptomic data of *M. graminicola* in susceptible rice plants, encompassing all life stages of the nematode from pre-parasitic J2 to female. We performed dual RNA-seq, in which the simultaneous sequencing of a host and its pathogen is realized (Westermann *et al.*, 2012), and developed a bioinformatics strategy to describe the *de novo M. graminicola* transcriptome. In order to identify candidate effectors, we searched for putative secreted proteins in the predicted *M. graminicola* proteome and exploited expression data to select genes that were up-regulated in the early parasitic stages of the interaction.

RESULTS

Meloidogyne graminicola RNA-seq data

RNA-seq was performed on 14 cDNA libraries originating from two independent biological assays (A and B) of seven samples each (Table 1). Two samples were freshly hatched pre-parasitic J2 nematodes. Ten samples were rice root tips collected at 2, 4, 8, 12 and 16 days after *M. graminicola* inoculation. Histological analyses showed that these time points cover the nematode life cycle from the parasitic J2 stage (2–8 dpi) to the J3/J4 stages (12 dpi) and the adult stage (16 dpi), just before egg-laying (Nguyen *et al.*, 2014). Lastly, root tips from mock-inoculated plants were collected at the same time point as the '2-dpi' time point to be used as control plant samples, especially as negative controls for gene expression analyses. For the nematode-infected rice samples, we chose to perform dual RNA-seq to avoid separating the nematodes from the rice roots. This delicate step, including enzymatic methods and mechanical manipulations, may have side-effects on the pathogen and may alter the natural state of the nematode

Table 1 Description of *Meloidogyne graminicola* RNA-sequencing (RNA-seq) libraries.

Biological assay	Library name	Inoculation time point	*M. graminicola* stage (main stage)	Read pairs	GC (%)
A	J2	–	Pre-parasitic J2	34 024 074	37
B	J2	–	Pre-parasitic J2	31 774 556	37
A	Nip	Mock '2 dpi'	–	32 551 940	48
A	Nip2	2 dpi	Parasitic J2	32 263 267	48
A	Nip4	4 dpi	Parasitic J2	30 335 264	49
A	Nip8	8 dpi	Parasitic J2	35 015 567	48
A	Nip12	12 dpi	J3–J4	40 213 348	43
A	Nip16	16 dpi	Female	35 749 433	38
B	Nip	Mock '2 dpi'	–	35 498 167	49
B	Nip2	2 dpi	Parasitic J2	36 026 736	50
B	Nip4	4 dpi	Parasitic J2	35 280 340	49
B	Nip8	8 dpi	Parasitic J2	36 776 218	48
B	Nip12	12 dpi	J3–J4	38 515 828	46
B	Nip16	16 dpi	Female	41 653 116	39

dpi, days post-infection; Nip, *Oryza sativa* (Nipponbare) rice plants.

transcriptome. The high sequencing depth of RNA-seq technology may compensate for this extraction step and allow for the detection of pathogen sequences among plant sequences. Illumina technology generating 100-bp read pairs was employed to sequence the 14 cDNA libraries. Between approximately 30 and 41 million read pairs were obtained for the different libraries (Table 1). The GC content of individual libraries revealed differences among samples, with 48%–49% GC for the control rice samples and 37% GC for the *M. graminicola* pre-parasitic J2 samples. Intermediate GC values were obtained for reads from the mixed samples at 12 and 16 dpi (Table 1), reflecting the respective amounts of plant and nematode reads in the libraries depending on nematode development, although the nematode reads appeared to be overestimated at 16 dpi.

Overview of the bioinformatics strategy

A schematic description of the bioinformatics strategy is presented in Fig. 2. To establish the *M. graminicola* transcriptome in the absence of a reference genome, the strategy was first based on read mapping against the rice genome (Kawahara *et al.*, 2013). Unmapped reads, putatively originating from the nematode or other contaminating organisms, were recovered. They were added to the pre-parasitic J2 reads and used to perform a *de novo*

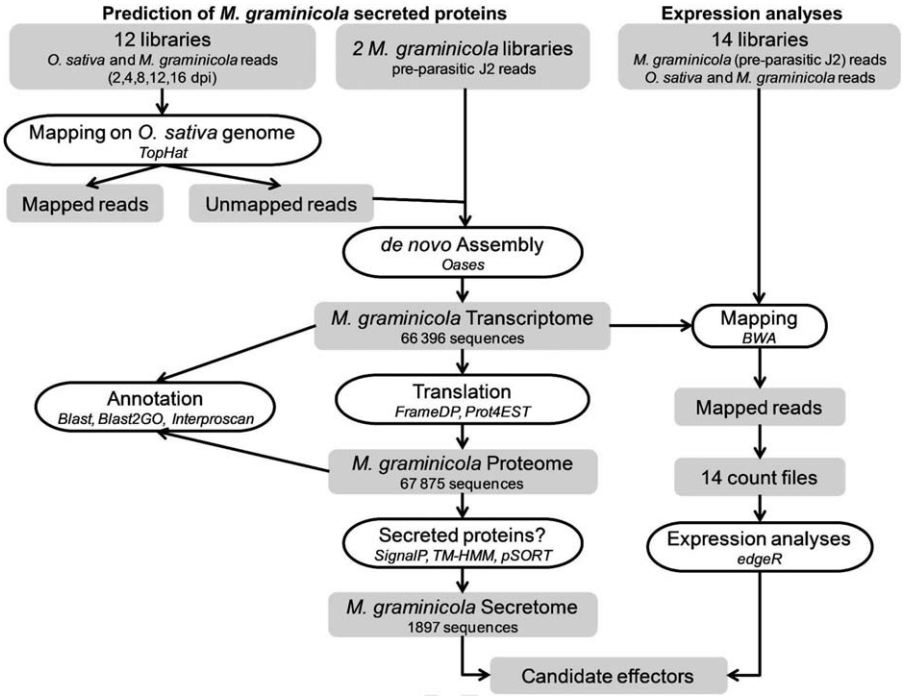

Fig. 2 Workflow summary of the bioinformatics strategy used for the identification of *Meloidogyne graminicola* candidate effectors. Bioinformatics tools used for each step are indicated in italic.

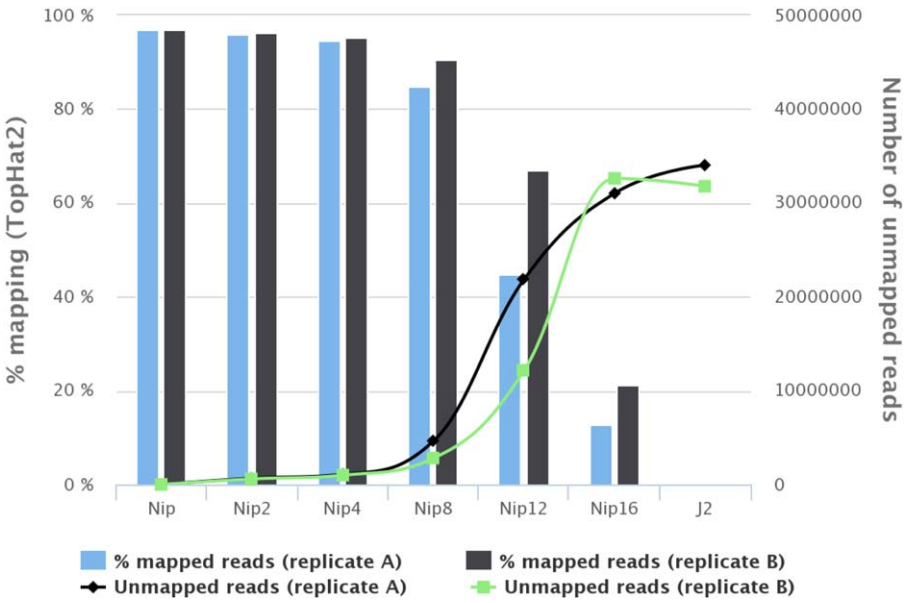

Fig. 3 Mapped and unmapped reads from the 14 RNA-sequencing (RNA-seq) libraries on the *Oryza sativa* genome. The *x*-axis shows the library names (see Table 1 for description), with the percentage of mapped reads on the left scale and the number of unmapped read pairs that were consequently used to perform the *de novo* assembly step (except unmapped read pairs from the rice control sample) on the right scale.

assembly of the *M. graminicola* transcriptome. Protein sequences were then predicted from the assembled sequences and putative secreted proteins were searched. Transcripts and protein annotations were performed to assign a function to each transcript. In parallel, gene expression levels were analysed in order to identify genes differentially expressed during the nematode life cycle. For those transcripts encoding putative secreted proteins, expression patterns helped to select for candidate effectors.

Separation of rice and nematode reads

All reads obtained from *O. sativa* samples were mapped to the rice Nipponbare genome using the TopHat program, with a GTF (Gene Transfer File) annotation file supplied as an option. The percentage of mapping obtained for each library gradually diminished from approximately 97% for the control plants to approximately 13%–21% for the nematode-infected plants at 16 dpi, depending on the biological replicate (Fig. 3). The respective quantities of rice sequences during the time course analysis reflected the development of the nematode inside roots, with a probable larger number of nematodes in the biological replicate A, or a faster development of the nematodes in this replicate. As observed for the GC (%) values, nematode reads seemed to be overestimated at 12 and 16 dpi. Pre-parasitic J2 reads were also mapped onto the rice genome to serve as negative control, resulting in less than 0.1% of mapped reads.

De novo assembly of the *M. graminicola* transcriptome

Reads unmapped to the rice genome and pre-parasitic J2 reads were further pooled, yielding a total number of 174 264 051 read pairs. These reads were used to perform the assembly step employing Oases, a specific software program that allows *de novo* RNA-seq assembly in the absence of a reference genome (Schulz *et al.*, 2012). After elimination of redundancy and sequences below 150 bp, 87 878 transcripts were assembled (Table 2). To eliminate potential ribosomal and rice remaining sequences, cleaning was performed with Seqclean, leading to a reduced number of 76 228 transcripts. Sequences of potential rice-contaminating micro-organisms were eliminated by comparing RPKM (reads per kilobase per million mapped reads) values in control plants with RPKM values in infected plants, leading to a final number of 66 396 transcripts. These transcripts were organized into 51 825 loci, as defined by Oases, including 41 511 loci containing one transcript and 10 314 loci containing at least two transcripts. The mean and median sequence sizes are 920 and

Table 2 Summary statistics for *Meloidogyne graminicola* sequence assembly and cleaning.

Read pairs used for *de novo* assembly	174 264 051
Assembled sequences (Oases)	87 878
Seqclean (rRNA)	851
Seqclean (*O. sativa*)	10 799
Cleaning (RPKM Nip > RPKM others)	9832
No. of transcripts after cleaning	66 396
No. of loci	51 825
No. of loci with 1 transcript	41 511
No. of loci with 2 or more transcripts	10 314
Length statistics	
Mean size (bp)	920
Median size (bp)	467
Longest transcript size (bp)	19 762
No. of transcripts ≥ 1 kb	18 633

RPKM, reads per kilobase per million mapped reads.

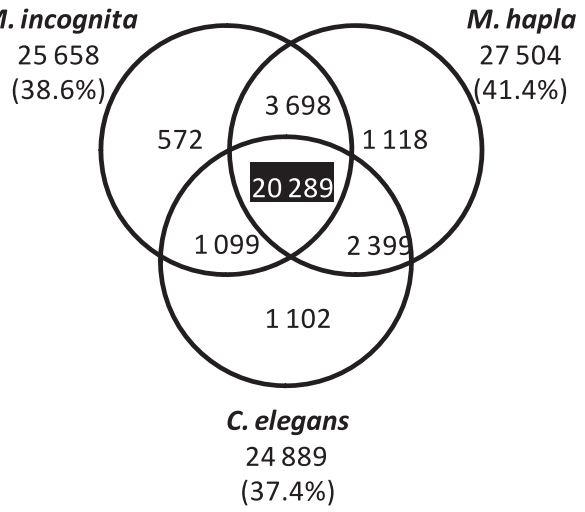

Fig. 4 BLASTX analyses of the *de novo Meloidogyne graminicola* transcriptome. Venn diagrams show the number and percentage of significant hits detected by BLASTX searches of the 66 396 *M. graminicola* transcripts against nematode (*M. incognita*, *M. hapla* and *Caenorhabditis elegans*) protein databases.

467 bp, respectively. The longest transcript size is 19 762 bp and around 28% of the transcripts are longer than 1 kb (Table 2).

Functional annotation of the *M. graminicola* transcriptome

BLASTX performed on the 66 396 assembled sequences against the UniProt database (which does not contain all *Meloidogyne* sequences) indicated that *M. graminicola* transcripts showed similarities with 41.8% of the Uniprot proteins. When testing against nematode protein databases from the model nematode *Caenorhabditis elegans* (26 150 proteins), and from the two RKNs sequenced and annotated, *M. incognita* (20 359 proteins) and *M. hapla* (14 420 proteins) (Abad *et al.*, 2008; Opperman *et al.*, 2008), significant hits were found for 45.6% of the *M. graminicola* transcripts. Most of the similarities were found with *M. hapla* (41.4%), followed by *M. incognita* (38.6%) and *C. elegans* (37.4%) (Fig. 4). Thus, combined BLASTX results indicated that only 52.1% of the *M. graminicola* sequences had similarities with recorded proteins from international databases. To determine which proportions of *M. hapla* and *M. incognita* annotated genes are covered by the *M. graminicola* sequences, TBLASTN was performed on the *M. hapla* and *M. incognita* protein sequences against the *M. graminicola* transcripts. Significant hits were found for 11 011 (76.4%) and 14 875 (73.1%) *M. hapla* and *M. incognita* proteins, respectively.

BLAST2GO analyses performed on the *M. graminicola* transcripts indicated that 37.9% of the sequences had at least one annotation characterized by Gene Ontology (GO) terms. Occurrences of GO terms were retrieved and classified according to the three BLAST2GO categories: 'Biological Process', 'Cell Compartment' and 'Molecular Function'. The distribution of *M. graminicola* transcripts across the GO categories was established (Fig. S1, see Supporting Information). For 'Biological Process' terms, genes associated with cellular and metabolic processes were found to be most abundant, followed by single-organism processes and biological regulation. For 'Molecular Function' terms, categories such as binding, catalytic, structural molecule and transporter activities were highly represented.

A total of 67 875 *M. graminicola* protein sequences were predicted from the assembled transcripts; this was more than the initial transcript number as FrameDP software is able to detect several open reading frames (ORFs) in a transcript. The search of functional domains conducted on the predicted proteins using the InterProScan program returned significant hits for 52.7% of the predicted proteins in the InterProScan database and 47.8% in the Pfam database. Data were classified according to the relative abundance of each functional domain and revealed the prevalence of domains involved in multiple cellular processes (Table S1, see Supporting Information). Protein kinases (2.08% of Pfam domains) are regulators of almost all biochemical pathways and direct most cellular processes, particularly in signal transduction. P-loop nucleoside triphosphate hydrolases (2.62% of InterProScan domains) may be involved in many molecular processes, including the induction of conformational changes in other molecules. Proteins containing WD domains (1.30% of Pfam domains) also play a central role in many biological processes, despite having no known intrinsic catalytic activity. They mainly provide platforms to assemble proteins or nucleic acids into functional complexes. The abundance of similar protein families has been reported for other parasitic nematodes, such as the rice pathogen *Hirschmanniella oryzae* (Bauters *et al.*, 2014).

Prediction of *M. graminicola* secreted proteins

Based on their protein sequences, putative secreted proteins were predicted according to the two following criteria: the presence of a signal peptide at the N-terminal region and the absence of a transmembrane domain. As a result, 1897 sequences (2.8% of total predicted proteins) were identified, ranging from 28 to 1821 amino acids (or 10 to 1802 amino acids when considering mature proteins) (Table S2, see Supporting Information). Around 45% (867) of the mature proteins were less than 100 amino acids. The predicted secretome was analysed in order to identify cysteine (Cys)-rich proteins and putative nuclear proteins. The results showed that 152 sequences are highly enriched in Cys (Cys content $>$ 8%) and 109 sequences have a putative nuclear localization signal (NLS) (pSORTII and NucPred predictions having nuclear scores of $>$50% and $>$0.5, respectively) (Table S2). At least one annotation (BLAST hit or InterProScan domain) was found for 55% of the sequences. We compared the relative abundance of the

functional domains identified in the predicted secreted proteins with regard to total proteins (Fig. 5, Table S1). This analysis essentially revealed enrichment for putative peptidases, proteinase inhibitors, thioredoxins and glycoside hydrolases in the predicted *M. graminicola* secretome, typical of proteins functioning outside the cell. They may be essential proteins for nematode development (for example, digestive proteases); however, many secreted proteins belonging to these classes are also known to be involved in parasitism and may participate in the degradation of host tissues (peptidases, glycoside hydrolases), detoxification of the environment (thioredoxins) or suppression of host immunity (proteinase inhibitors), as described previously in plant- and animal-parasitic helminths (Haegeman *et al.*, 2012; Hewitson *et al.*, 2009).

Differential gene expression across the *M. graminicola* life cycle

Transcripts presenting variable abundance levels during the pathogen life cycle could play a role in nematode development, but also in parasitism. Expression levels of *M. graminicola* genes across its life cycle were determined by mapping each library against the *de novo M. graminicola* transcriptome (Fig. S2, see Supporting Information). Differential expression analyses were then performed between successive time points based on the count files generated by the mapping step with the dedicated software EdgeR. The numbers of up- and down-regulated genes were determined for each comparison (Table 3). Overall, major modifications in gene expression were detected at 8 dpi for up-regulated genes (343 induced genes) and at 12 dpi for down-regulated genes (140 repressed genes). BLASTX hits, InterProScan annotations and RPKM values were recovered for the differentially expressed genes (Table S3, see Supporting Information). An annotation was found for 37% of the genes, allowing the identification of different categories of regulated genes during the *M. graminicola* life cycle.

In the first steps of root invasion, nematodes need to degrade plant cell walls to facilitate penetration and migration. Two *M. graminicola* genes encoding endoglucanases, specific for cellulose, the major component of the plant cell wall, and one gene with a glycoside hydrolase (GH) signature were induced at 2 dpi. Two other genes encoding cell wall-degrading enzymes, one poly-α-D-galacturonosidase and one glucuronoxylanase, were over-expressed during the pre-parasitic J2 stage and at the early parasitic stages. The down-regulation of these genes at 4 dpi may be correlated with the end of *M. graminicola* migration inside the host, once the nematode enters the sedentary phase.

After the initiation of a permanent feeding site, parasitic J2 nematodes feed and grow. They become swollen and evolve to J3, J4 and adult stages through three successive moults. As early as 4 dpi, but mostly at 8 dpi, numerous genes related to the nematode moulting process were induced. More than 40 genes encoding cuticle collagen proteins or cuticlin proteins, the structural proteins that make up the extracellular envelop of the nematode body, were up-regulated. In parallel, eight genes encoding *nas* zinc-metalloproteinases, enzymes belonging to a peptidase family known to be involved in the shedding of the old cuticle, one gene encoding a MLT-7 peroxidase, involved in collagen cross-linking, and five genes encoding MLT-10-type proteins, involved in the moulting cycle (Page *et al.*, 2014), were induced. The activation of essential genes related to translation (elongation factors and

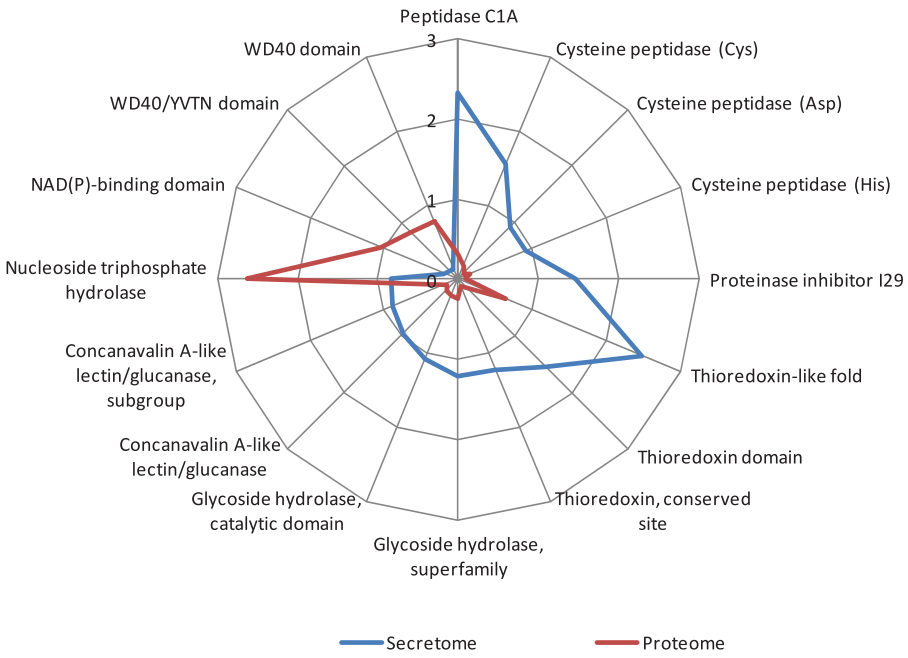

Fig. 5 Relative abundance of functional domains in the predicted secretome compared with the whole proteome. The Kiviat diagram shows the frequency (per cent) of InterProScan (IPR) domains in the proteome and secretome. The 16 functional domains having the highest differences in frequency between the two datasets are presented.

Table 3 Differential gene expression across the *Meloidogyne graminicola* life cycle in rice plants.

Compared samples (EdgeR)	2 dpi vs. pre-J2	4 dpi vs. 2 dpi	8 dpi vs. 4 dpi	12 dpi vs. 8 dpi	16 dpi vs. 12 dpi
Up-regulated genes	21	59	343	99	109
Down-regulated genes	3	19	74	140	89

Numbers of up-regulated and down-regulated genes were determined by EdgeR analyses between two successive samples of the *M. graminicola* infection time course. dpi, days post-infection.

ribosomal proteins) was associated with moulting genes regulation and traduced an intense metabolic step. All of these genes were transiently expressed and their expression was down-regulated at 12 dpi, once moulting had occurred (Nguyen *et al.*, 2014; Fig. 1).

At the adult stage, genes related to reproduction and early embryonic development were up-regulated. Two genes encoding proteins with nematode major sperm protein (MSP) signatures and one gene corresponding to a *C. elegans* germ cell-expressed protein were induced at 12 dpi. At 16 dpi, two homeobox genes involved in the regulation of anatomical development patterns were induced. These data may reflect the induction process of embryonic larval development and may be correlated with the observation of *M. graminicola* egg-laying adult females in the subsequent days (18–22 dpi) (Nguyen *et al.*, 2014; Fig. 1).

At all infection stages, genes encoding peptidases were induced in the *M. graminicola* transcriptome. Important biological roles could be attributed to this class of enzymes. Basically, peptidases may digest host proteins during feeding to serve as a source of nutrients. They may also play an important role in parasitism, either by degrading plant cell wall proteins to facilitate migration, or by degrading plant defence proteins (Haegeman *et al.*, 2012). *Meloidogyne graminicola* genes encoding proteases, such as an aspartic peptidase, a SYS1 homologue and a legumain protease, were up-regulated in the early parasitic stages. A cathepsin-cysteine proteinase-encoding gene, induced from 4 to 16 dpi, may play a role in all sedentary stages.

Identification of candidate effector genes

To identify *M. graminicola* putative effector genes, we focused on the putative secreted protein genes up-regulated in the early steps of the interaction with the host plant (especially at 2 and 4 dpi), with RPKM values >50 and without functions linked to general metabolism or developmental processes. Of the 1897 predicted secreted proteins, 15 up-regulated candidate genes were selected to conduct further analyses (Table 4). The cDNA sequences were successfully cloned and Sanger sequences confirmed the *de novo* assembled sequences (data not shown). The 15 deduced sequences have a small size (64-386 amino acids). However, two sequences have no stop codons and are supposed to be partial (Table 4). Four proteins (Mg40980, Mg12322, Mg13532, Mg28330) have a high Cys content (>8%) and two others may be addressed to the nucleus (Mg17098, Mg16593). Only four of the 15 candidate proteins have predicted functions. Mg40980 encodes a Cys-rich protein with a metallothionein domain. Mg4965, although probably missing the C-terminal part, shows similarities with *Meloidogyne*

Table 4 Description of the 15 candidate effectors identified in this study.

Gene	ORF (bp)	Protein (amino acids)	Cys (no./%)	NLS*	Highest expression†	Predicted domains
Mg449	405	134	2/1.5	N	J2	Carbohydrate-binding domain
Mg757	219	72	0/0	N	J2 – 2 dpi	–
Mg4610	864	287	2/0.7	N	J2 – 4 dpi	–
Mg40980	264	87	12/13.8	N	2 dpi	Metallothionein
Mg4965	573	191p	1/0.5	N	2 dpi	CLE-like
Mg12322	363	120	19/15.8	N	2 dpi	–
Mg12162	818	273p	3/1.1	N	2 dpi	–
Mg17272	519	170	2/1.2	N	2 dpi	–
Mg17098	351	116	1/0.9	Y	2 dpi	–
Mg28330	195	64	9/14.1	N	2 dpi	–
Mg46375	312	103	1/1	N	2 dpi	–
Mg16593	1161	386	3/0.8	Y	4 dpi	–
Mg66296	345	114	1/0.9	N	4 dpi	–
Mg13532	246	81	6/7.4	N	2–16 dpi	–
Mg11937	744	247	11/4.5	N	J2, 16 dpi	VAP1

Cys, cysteine; dpi, days post-infection; ORF, open reading frame; p, partial.

*NLS, nuclear localization signal prediction (Y, yes; N, no).

†Highest expression based on RPKM (reads per kilobase per million mapped reads) values.

proteins containing plant ligand-like motifs (CLE-like motifs) (Rutter *et al.*, 2014b). Its sequence is structured with at least nine CLE-like motifs, composed of 16 amino acids each, including the core consensus sequence 'YPQGPE' (Fig. S3, see Supporting Information). Mg11937 shows similarities to the GrVAP1 effector from *Globodera rostochiensis* (Lozano-Torres *et al.*, 2012, 2014). Lastly, Mg449 exhibits a carbohydrate-binding domain.

To validate the RNA-seq data, quantitative reverse transcription-polymerase chain reaction (RT-qPCR) experiments were performed using three independent biological replicates of pre-parasitic J2 and parasitic *M. graminicola*, the two replicates used for RNA-seq, plus a third. Expression profiles were successfully obtained for 10 of 15 candidate effector genes (Fig. 6). *Mg449* and *Mg757* were highly expressed at the pre-parasitic J2 stage. Transcript levels of *Mg40980*, *Mg4965*, *Mg12322*, *Mg17272, Mg16593* and *Mg66296* showed a steady increase at 2 or 4 dpi. *Mg13532* showed continual transcript accumulation from 2 to 16 dpi, and *Mg11937* was up-regulated at the pre-parasitic J2 stage and at 16 dpi. Highly consistent results were obtained when comparing qPCR gene expression levels with RPKM values, indicating a good correspondence between RT-qPCR and RNA-seq methodologies (Fig. S4, see Supporting Information).

In situ hybridizations were performed to localize the expression of six candidate genes in the nematode body. Pre-J2 nematodes were used for genes highly expressed at this stage (Mg449 and Mg4965). A mixture of pre-J2 and parasitic nematodes excised from host roots at 4 or 8 dpi was used for the other genes. The expression of four of the six genes tested was detected by immunostaining of the digoxigenin (DIG)-labelled probes (Fig. 7). The antisense probes against the *Mg449* and *Mg4965* genes strongly stained the subventral gland regions from pre-J2s. The *Mg16593* probes stained the dorsal gland area of parasitic nematodes collected at 4 dpi and gave no signal on pre-J2s. The *Mg13532* probes stained the intestinal area of both pre-J2 and parasitic nematodes collected at 8 dpi. The *Mg40980* and *Mg12322* probes returned no detectable signal on pre-J2 or parasitic nematodes collected at 4 dpi. All sense probes displayed no visible signal.

DISCUSSION

Because of its damage to rice crops, *M. graminicola* is an important pathogen that requires greater biological and molecular characterization in order to develop novel control strategies. In the absence of a reference genome, dual RNA-seq represents a good opportunity to obtain a large amount of data on the nematode genes expressed during the biotrophic interaction with rice. The simultaneous sequencing of plant and pathogen mRNAs has been successfully applied to decipher molecular plant–pathogen interactions (Kawahara *et al.*, 2012; Teixeira *et al.*, 2014; Yang *et al.*, 2013) even without the reference genome (Fernandez *et al.*, 2012; Thakur *et al.*, 2013; Tremblay *et al.*, 2013). In this study, 12 transcriptomic datasets were generated from two biological replicates of six different *M. graminicola* life stages in rice plants from pre-parasitic J2 to adult female stages. These data produced a comprehensive *M. graminicola* transcriptome, representing an important resource for our understanding of the biology of this parasitic nematode in unprecedented detail.

Starting from 176 million read pairs, the Oases assembler predicted 66 396 transcripts that were organized in 51 825 loci (Table 2). Further cloning and Sanger sequencing of a set of 15 genes suggested an overall good assembly of the reads. However, BLASTX results indicated that some assembled transcripts may contain two ORFs, or even three, which correspond to juxtaposed genes in the *M. incognita* genome (data not shown). Such transcripts may result from *in silico* sequence mis-assembly of genes having overlapping regions in compact genomes (Matin and Wang, 2011), such as for *M. graminicola*, which has one of the smallest animal genomes estimated so far (30 Mb) (Gregory *et al.*, 2007). Functional annotations indicated that only 52% of the *M. graminicola* transcripts returned significant hits in protein databases (Uniprot or nematode species) or had predicted domains in their protein sequences. More specifically, similarities were found with 75% of the proteins of other RKN species (11 060 *M. hapla* proteins and 14 911 *M. incognita* proteins). Almost half of the *M. graminicola* genes had no annotation, even including some highly expressed genes, as defined by RPKM values. These genes may encode novel proteins with primary structures too divergent from those of other nematode species to find homologues in the international databases, and may reflect an evolutionary adaptation to the specific *M. graminicola* life style.

Gene expression data obtained at several *M. graminicola* life stages allowed us to decipher the nematode transcriptional patterns during rice root infection (Table S3). Essential genes for nematode development were clearly identified and their temporal expression profiles corresponded to the nematode developmental stages observed by histological analysis (Nguyen *et al.*, 2014). For instance, cuticle biosynthesis genes involved in successive moulting to J3, J4 and adult stages were highly expressed from 8 to 12 dpi, and were less expressed around 16 dpi when females had developed (Table S3). Genes related to parasitism were also highlighted, such as cell wall-degrading enzymes (glycoside hydrolases) expressed until 4 dpi, which may play a role in root invasion, or proteases, which may act at multiple parasitism steps. We also identified some differentially expressed genes that had no predicted function or homologues in the databases. Their expression profiles may provide some clues about their role in *M. graminicola* biology. For instance, more than 40 genes that specifically displayed high expression (RPKM $>$ 100) at 16 dpi may be associated with the adult female stage or be involved in egg development.

To identify putative proteins involved in the *M. graminicola* infection process, we focused on putatively secreted proteins and

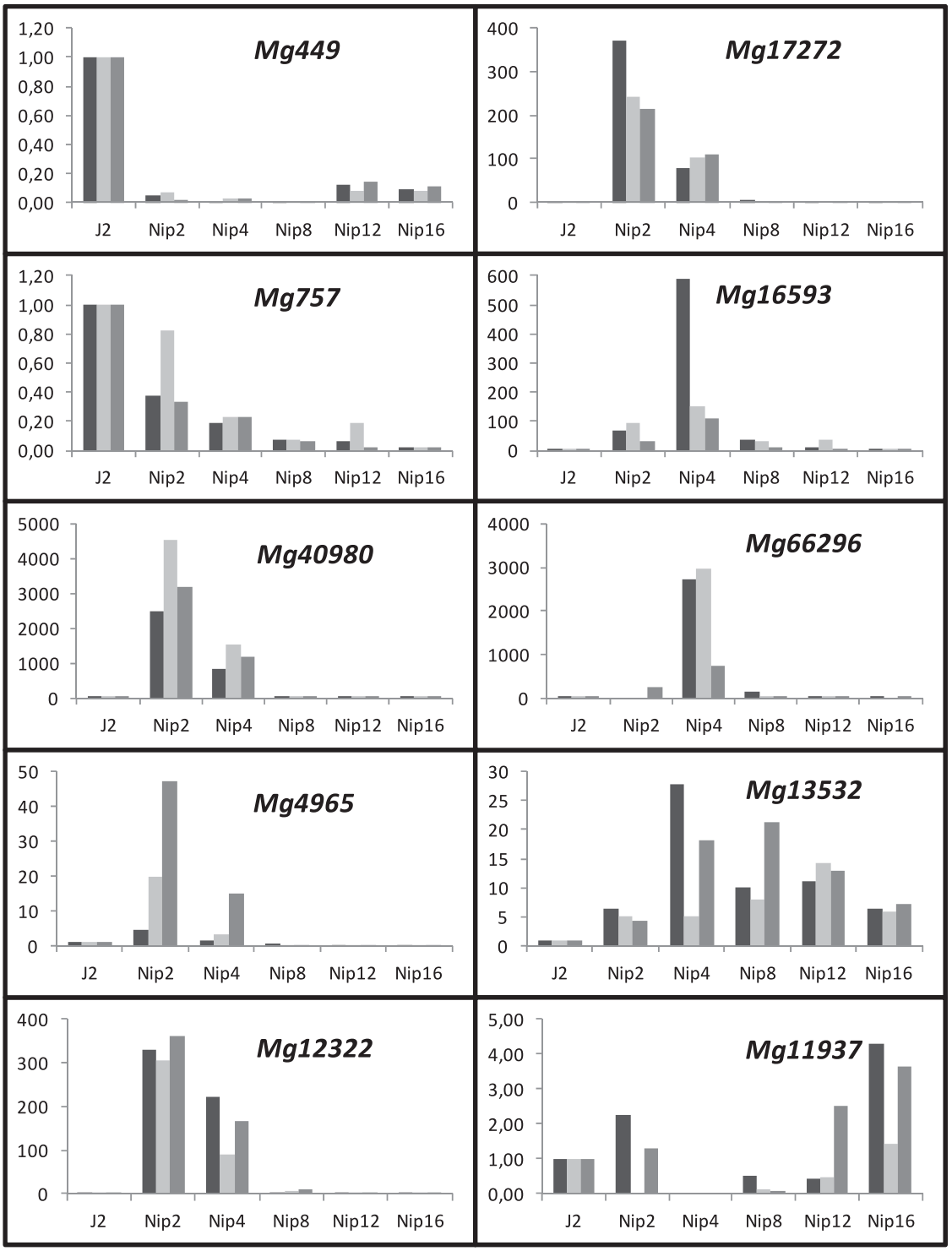

Fig. 6 Expression profiles of *Meloidogyne graminicola* candidate effector genes determined by quantitative reverse transcription-polymerase chain reaction (RT-qPCR) analyses. *y* axes represent the relative fold change in gene expression levels in various *M. graminicola* parasitic stages in rice plants compared with the pre-parasitic J2 stage. *Mg-actin* was used as an internal control gene for normalization. Three independent biological assays were analysed: dark grey, replicate A; light grey, replicate B [both used for RNA-sequencing (RNA-seq) libraries]; mid-grey, replicate C.

Fig. 7 Localization of *Meloidogyne graminicola* gene expression by *in situ* hybridization. Digoxigenin-labelled antisense probes of candidate genes were hybridized to transcripts expressed within the nematode body at the pre-J2 stage (*Mg449* and *Mg4965*) and parasitic stages [4 days post-infection (dpi) for *Mg16593* and 8 dpi for *Mg13532*] of *M. graminicola*. Dark-stained signals are indicated by arrows; M indicates the metacorpus.

combined a 'secreted protein prediction' pipeline with differential expression analyses based on RNA-seq data. Based on the hypothesis that genes up-regulated in the early steps of the infection could play a role in nematode parasitism, we selected a set of 15 candidate genes from the 1897 predicted secreted proteins found in the *M. graminicola* transcriptome (Table 4). Expression profiles further established by RT-qPCR assays on 10 genes validated the RPKM values, even in the 2-, 4- and 8-dpi libraries in which nematode transcripts are represented by a low read number (less than 15%) (Fig. 6).

Only four *M. graminicola* candidate genes have an identified functional domain. The *Mg40980* gene encoding a metallothionein is highly up-regulated at 2 and 4 dpi (Fig. 6). Metallothioneins are low-molecular-weight and Cys-rich proteins, involved in heavy metal detoxification and homeostasis. Moreover, they may play a role in cell protection following the exposure to reactive oxygen species (ROS) (Ruttkay-Nedecky *et al.*, 2013; Zeitoun-Ghandour *et al.*, 2011). *Mg40980* may be an important component of antioxidant proteins known to be produced by plant-parasitic nematodes to scavenge host ROS species (Haegeman *et al.*, 2012). Two other 2-dpi up-regulated genes, *Mg12322* and *Mg28330*, also encode Cys-rich proteins and lack sequence homologues in databases. These proteins usually have structural homologies, based on Cys residues, forming disulfide bonds that play a critical role in folding and, consequently, the chemical stability of the molecule. Although small and Cys-rich proteins are a hallmark of fungal and oomycete effectors (Stergiopoulos and de Wit, 2009), they have never been described as characteristic effectors in nematodes, but may be important players in the molecular interactions between the two partners.

Mg4965 is another interesting *M. graminicola* candidate induced at 2 dpi, when feeding sites are initiated (Nguyen *et al.*, 2014). The Mg4965 protein contains multiple repetitions of a 16-amino-acid motif similar to plant CLE peptides (Miyawaki *et al.*, 2013). CLE genes are a large family of plant genes involved in cell-to-cell communication, with CLAVATA3 being the best studied. From the mature proteins, CLE motifs are processed into 12- or 13-amino-acid peptides that act as ligands for receptor-like kinases and induce developmental changes within plant cells (Miyawaki *et al.*, 2013). It has been discovered that CLE-like proteins are conserved in cyst nematodes and RKNs (Miyawaki *et al.*, 2013; Rutter *et al.*, 2014b). Cyst nematode CLE peptides have been shown to functionally mimic the effects of plant CLE peptides and probably trigger developmental changes necessary for feeding site formation (Mitchum *et al.*, 2012). A recent study has demonstrated that the potato cyst nematode protein GrCLE1 is cleaved and glycosylated *in planta* before interacting with a potato CLAVATA2-like receptor (Chen *et al.*, 2014). The Mg4965 CLE-like motifs exhibit the typical proline residue conservation of plant CLE peptides and may be processed by the host cellular machinery, like GrCLE1, to promote parasitism. Remarkably, Mg4965 has at least nine CLE-like motifs and the mature protein may yield multiple peptides, probably to amplify the CLE signal. Cyst nematode CLE genes are specifically expressed in the dorsal gland and during all parasitic life stages of the nematode (Lu *et al.*, 2009; Wang *et al.*, 2001). Conversely, *Mg4965* is expressed in the subventral glands of *M. graminicola* pre-J2s (Fig. 7) and is induced until 4 dpi (Fig. 6), suggesting a role in the early steps of parasitism.

We selected the *Mg11937* gene, encoding a venom allergen-like protein (VAP) identical to the *M. graminicola contig06883* previously identified and specifically expressed in the subventral glands of pre-J2s (Haegeman *et al.*, 2012). The VAP proteins from two cyst nematodes, *G. rostochiensis* (GrVAP1) and *Heterodera schachtii* (HsVAP1 and HsVAP2), have been shown recently to be important molecular players in the suppression of plant defences (Lozano-Torres *et al.*, 2014). Ectopic expression of GrVAP1, HsVAP1 and HsVAP2 resulted in the loss of basal immunity to multiple pathogens in Arabidopsis. Conversely, knocking down the expression of *GrVAP1* severely affected the infectivity of *G. rostochiensis*. The modulation of basal immunity by VAP proteins involves extracellular

proteases. Hence, GrVAP1 targets the apoplastic papain-like cysteine protease $Rcr3^{pim}$ of tomato and activates *Cladosporium fulvum* 2 (Cf-2)-mediated nematode resistance (Lozano-Torres *et al*., 2012). It would be very interesting to determine whether *Mg11937* has a similar function in rice–nematode interactions.

No specific function was identified for the 11 other selected genes. However, prediction tools revealed an NLS in the Mg16593 protein, suggesting that it could be targeted to the plant cell nucleus, like other plant-parasitic nematode effectors (Quentin *et al*., 2013). Two RKN effectors, Mi-EFF1 and Mj-NULG1, are immunolocalized within giant cell nuclei and seem to be specific to the early steps of parasitism (Jaouannet *et al*., 2012; Lin *et al*., 2013). Up-regulation of the *Mg16593* gene at 2 and 4 dpi indicates that it may also play a role in the early steps of interaction with rice plants. Moreover, its specific expression in the dorsal gland suggests that the protein is secreted through the stylet into the host tissues. It remains to be determined whether Mg16593 is imported into the host cell nuclei during *M. graminicola* parasitism. Another candidate gene, *Mg13532*, is expressed throughout the sedentary stage of *M. graminicola*. Genes with such continual expression may have important functions in nematode biotrophy. Its expression seems to be located in the intestinal tract of the nematode body and its role remains to be elucidated. Other candidate genes do not resemble any referenced proteins and could be considered as pioneers, as with many candidate effectors from plant-parasitic nematodes (Haegeman *et al*., 2012). The challenge is thus to determine their function and their contribution in parasitism.

In conclusion, in the absence of a reference genome for *M. graminicola*, dual RNA-seq allowed us to successfully establish a *de novo* transcriptome covering the J2 and parasitic *M. graminicola* stages, which complements the 454-transcriptome established on pre-parasitic J2s by Haegeman *et al*. (2013). The description of gene expression changes during the *M. graminicola* life cycle revealed key genes involved in nematode development, and provided insights into genes putatively associated with parasitism. We identified novel *M. graminicola* candidate effectors, such as CLE-like proteins, VAP and proteins involved in detoxification. Further research is now needed to specify their role in *M. graminicola* parasitism. Our study used bioinformatics approaches to predict secreted proteins on the basis of signal peptide sequences. It is now recognized that parasitic nematodes also release proteins that are not encoded by a signal peptide. Cell–cell communication may be mediated through the active secretion of small extracellular vesicles, called exosomes, that are taken up by host cells (Marcilla *et al*., 2012). Exosome-mediated transfer of small RNAs has also been demonstrated recently in animal-parasitic nematodes to modulate host immunity (Buck *et al*., 2014). It is thus probable that the proteins described here will only represent a fraction of the full effector repertoire used by *M. graminicola* to subvert its host plant.

EXPERIMENTAL PROCEDURES

Biological material and inoculation assays on rice plants

The *M. graminicola* population was originally collected from Laurel (Batangas, Philippines) and cultured on *O. sativa* cv. IR64. *Meloidogyne graminicola* pre-parasitic J2s were obtained as described previously (Nguyen *et al*., 2014). *Oryza sativa* (cv. Nipponbare) plants were inoculated with 100 J2s in plastic tubes containing synthetic absorbent polymer (SAP) substrate as described previously (Nguyen *et al*., 2014). At 1 dpi, plants were transferred to a 50-mL hydroponic culture system containing Hoagland solution to synchronize the infection process. Roots tips (1–2 mm) or visible galls were collected at 2, 4, 8, 12 and 16 dpi. Roots tips from non-inoculated plants were collected at the same time as the 2-dpi plants. Samples were immediately frozen in liquid nitrogen and stored at −80°C until further use. A mean of 25 plants was used for each sample. Three independent biological replicates were performed; two replicates were used for RNA-seq and three replicates were used for qPCR validation. Histological studies were performed as described previously (Nguyen *et al*., 2014).

RNA extraction and cDNA synthesis

Total RNA was extracted from pre-parasitic J2s or rice root samples using the RNeasy Plant mini kit (Qiagen, Courtaboeuf, France) with the addition of an on-column DNase I digestion. First-strand cDNAs were synthesized from 1 μg of total RNA in a final volume of 20 μL using an oligo-$dT_{(18)}$-MN primer (Eurogentec, Angers, France) and GoScript™ Reverse Transcriptase (Promega, Lyon, France) for cloning experiments or the Omniscript RT kit (Qiagen) for real-time qPCR experiments.

Construction of cDNA libraries and Illumina sequencing

RNA integrity was determined by the RNA Integrity Number using an Agilent, Santa Clara CA, USA RNA 6000 Nano Chip. RNA was quantified using the Quant-iT™ RiboGreen® RNA Assay Kit, and 2 μg were used per sample. Individually indexed libraries were prepared using the TruSeq RNA Sample Preparation v2 kit according to the manufacturer's instructions (Illumina, San Diego CA, USA). Each cDNA library was verified and quantified using the Agilent DNA 7500 Chip. The libraries were sequenced at the Centre National de Séquençage (Genoscope, Evry, France) on a HiSeq2000 system (Illumina Inc.) generating 100-bp read pairs. Each indexed library represented one-fifth of one flowcell lane. Raw data were submitted to European nucleotide archive/short read archive (ENA/SRA) (accession number ERS715982).

Preprocessing and mapping of Illumina reads

Contaminating adapter and index sequences were removed from the raw sequences. Low-quality bases ($Q < 20$) were trimmed out from both the 5′ and 3′ ends of the reads. Sequences with undetermined bases were trimmed from the second N to the extremity of the reads. Reads less than 30 nucleotides were discarded and only read pairs were kept for subsequent analyses. The quality of each library was controlled with FastQC.

The first base of all reads was consequently eliminated by an in-house script. Then, the reads were aligned to the *O. sativa* reference genome (MSU7 release, http://rice.plantbiology.msu.edu/) using TopHat version 2.0.6 (Trapnell *et al.*, 2009). A GTF annotation file derived from the genome was supplied using the –G option. Standard parameters were used, except for the following: -r 100 -mate-std-dev 50 -i 30 -I 20000 -a 8 -m 1 –no-coverage-search -g 10 –bowtie-n -library-type fr-unstranded –microexon-search.

De novo assembly of the *M. graminicola* transcriptome

Unmapped read BAM files generated by TopHat were converted to Fastq files so that they could be taken as input for the assembly step. Using Deinterlacer and Interlacer components of the Galaxy environment, only mate reads were retrieved and used for the *de novo* assembly of the *M. graminicola* transcriptome. Oases version 0.2.08 (Schulz *et al.*, 2012) was used with k-mer = 51. To remove redundant transcripts, the assembled contigs were processed with CD-HIT-EST (Li and Godzik, 2006) with an identity threshold of 95%. Low-complexity sequences were masked using DustMasker and sequences less than 150 bp were discarded. Contaminating sequences were eliminated in two steps. The remaining rice sequences or ribosomal sequences were eliminated with SeqClean to which we provided the following potential contaminant sequences: SSU and LSU sequence files from the SILVA ribosomal RNA Database (Quast *et al.*, 2013) and nuclear, mitochondrial and chloroplast chromosomes as well as cDNA gene models from the rice genome annotation MSU7. Potential contaminating sequences originating from other micro-organisms associated with rice roots were eliminated when their RPKM value in one control root sample was superior to the RPKM values of infected root samples (see 'Differential expression analyses' section for RPKM determination).

Search of homologies and annotation

BLASTX alignments were performed on the *C. elegans* and *M. hapla* protein databases (downloaded from the WormBase website version WS236), the *M. incognita* protein database (MincV1A1 file downloaded from https://www6.inra.fr/meloidogyne_incognita/), as well as the UniProt and NR databases (downloaded in April 2013). BLASTN alignments were performed on the nt database. TBLASTN alignments were also performed with *M. hapla* and *M. incognita* protein sequences against the *M. graminicola* assembled sequences. An e-value threshold of $1e^{-10}$ was applied and the best hit was kept. BLAST2GO was then performed on the assembled transcriptome based on Uniprot BLASTX results.

Prediction of the derived proteome

Transcripts were first translated using the FrameDP program (Gouzy *et al.*, 2009) with WORMPEP as a reference protein database, a minimum peptide length of 30 amino acids and the GC3 parameter. Transcripts without prediction were translated using prot4EST software (Wasmuth and Blaxter, 2004) based on BLAST results against the Uniprot database. Predictions were pooled and redundancy was eliminated using CD-HIT with an identity threshold of 100%. InterProScan was performed on the whole predicted proteins to identify protein signatures referenced in the Pfam and InterProDomain databases.

Identification of putative secreted proteins

Putative secreted proteins were predicted using both SignalP 4.1 for the presence of signal peptide and TM-HMM for the absence of transmembrane domain (Petersen *et al.*, 2011). The mature protein file generated by SignalP was used to predict the presence of putative NLS in the sequences using PSORT-II (Nakai and Horton, 1999) and NucPred (Brameier *et al.*, 2007) programs. Cys content was calculated for each mature protein and all data were compiled into an Excel table. These steps were implemented as a 'secreted protein prediction' workflow in the GALAXY platform.

Differential expression analyses

All the libraries were mapped to the assembled transcriptome with BWA-ALN software (Li and Durbin, 2009) using default options, except for maximum of differences = 5, maximum number or fraction of gap opens = 2. BAM files were converted to count files using Samtools idxstat, and to RPKM files using an in-house script. Differential gene expression was determined on the count files using edgeR, a Bioconductor software package (Robinson *et al.*, 2010), by comparing two samples and benefitting from the two available replicates. Differential expression of a given transcript was analysed only if the sum of the counts in the four samples was superior to 10. Analyses were performed with the 'Generalized Linear Models' method and the 'common dispersion' as a dispersion estimate method. The false discovery rate (FDR) adjustment method was based on Benjamini and Hochberg. Genes were considered as induced or repressed only when the $\log_2$FC [log(fold change)] value was >2 or less than -2, respectively, with a logCPM [log(count per million)] value >3 and FDR <0.05. In addition, for each transcript, EdgeR results were combined with RPKM values in Excel files. Tables were manually examined in order to eliminate potential 'contaminants' (loci in which the RPKM value is superior in plant samples than in infected plant samples), redundant transcripts (originating from loci containing several transcripts, as defined by the Oases assembler) or false positives (divergent RPKM values between biological replicates).

Cloning of candidate genes

Each PCR was performed on the cDNA sample in which the candidate presented the highest RPKM value. The amplification was carried out in a reaction volume of 25 μL with Advantage®2 polymerase mix (Clontech, St-Germain-en-Laye, France) and 200 nM of gene-specific primers, designed from the selected sequences (Eurogentec). After electrophoresis separation on a 1.5% agarose gel, bands of interest were cut from the gel, cloned into the pGEM®-TEasy vector (Promega) and sequenced (Beckman Coulter Genomics, Takeley Essex, UK).

Real-time qPCR assays and data analysis

Specific primers were designed from the *M. graminicola* cloned sequences using Beacon Designer 7.0 software (Premier Biosoft International, Palo Alto CA, USA) (Table S4, see Supporting Information). RT-qPCR assays were performed on cDNA samples (diluted 1: 100) as described previously (Nguyen *et al.*, 2014). The *M. graminicola actin* gene, previously cloned in

our laboratory (V. P. Nguyen, unpublished data), was used as reference gene, and the fold change was calculated using the $2^{-\Delta\Delta Ct}$ method.

In situ hybridization experiments

In situ hybridization was performed on *M. graminicola* larvae, essentially as described by De Boer *et al.* (1998). DIG-labelled probes were synthesized in sense and antisense directions with specific forward and reverse primers (Table S4). Freshly hatched J2s (pre-J2s) were fixed in 2% paraformaldehyde for 16 h at 4°C, followed by 4 h at room temperature. Parasitic populations of nematodes were extracted from rice root galls by gentle grinding in 2% paraformaldehyde using a mortar and a pestle, followed by progressive sieving (250, 80 and 5 μm). These parasitic nematodes were supplemented with freshly hatched pre-J2s and fixed as described above. Nematode sections were hybridized overnight at 40 C with the sense or antisense probes. The bound probes were detected by alkaline phosphatase immunostaining (DIG Nucleic Acid Detection Kit, Roche, Meylan, France). The stained nematodes were visualized using a Zeiss, Marly-le-Roi, France, AXIO Imager inverted light microscope, and images were taken with a digital camera with standard bright-field optics.

ACKNOWLEDGEMENTS

We wish to thank Jamel Aribi (IRD) for help with the maintenance of rice plants, Georges Reversat (IRD) for advice on nematode systems and François Sabot for helpful discussions on the bioinformatics approach. This work was funded by CEA/Genoscope-INRA-IRD collaborative project N°9-2012 '*Meloidogyne graminicola* effectors'. The authors declare no conflicts of interest.

REFERENCES

Abad, P., Gouzy, J., Aury, J.M., Castagnone-Sereno, P., Danchin, E.G., Deleury, E., Perfus-Barbeoch, L., Anthouard, V., Artiguenave, F., Blok, V.C., Caillaud, M.C., Coutinho, P.M., Dasilva, C., De Luca, F., Deau, F., Esquibet., M, Flutre, T., Goldstone, J.V., Hamamouch, N., Hewezi, T., Jaillon, O., Jubin, C., Leonetti, P., Magliano, M., Maier, T.R., Markov, G.V., McVeigh, P., Pesole, G., Poulain, J., Robinson-Rechavi, M., Sallet, E., Ségurens, B., Steinbach, D., Tytgat, T., Ugarte, E., van Ghelder, C., Veronico, P., Baum, T.J., Blaxter, M., Bleve-Zacheo, T., Davis, E.L., Ewbank, J.J., Favery, B., Grenier, E., Henrissat, B., Jones, J.T., Laudet, V., Maule, A.G., Quesneville, H., Rosso, M.N., Schiex, T., Smant, G., Weissenbach, J. and Wincker, P. (2008) Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*. *Nat. Biotechnol.* **26**, 909–915.

Bauters, L., Haegeman, A., Kyndt, T. and Gheysen, G. (2014) Analysis of the transcriptome of *Hirschmanniella oryzae* to explore potential survival strategies and host–nematode interactions. *Mol. Plant Pathol.* **15**, 352–363.

Bellafiore, S., Shen, Z., Rosso, M.N., Abad, P., Shih, P. and Briggs, S.P. (2008) Direct identification of the *Meloidogyne incognita* secretome reveals proteins with host cell reprogramming potential. *PLoS Pathog.* **4**(10), e1000192.

Benjamini, Y. and Hochberg Y. (1995) Controlling the False Discovery Rate: A practical and powerful approach to multiple testing. *J. R. Statist. Soc. B* **57**, 289–300.

Brameier, M., Krings, A. and MacCallum, R.M. (2007) NucPred—predicting nuclear localization of proteins. *Bioinformatics*, **23**, 1159–1160.

Bridge, J., Plowright, R.A. and Peng, D. (2005) Nematode parasites of rice. In: *Plant-Parasitic Nematodes in Subtropical and Tropical Agriculture* (**Luc, M., Sikora, R.A.** and **Bridge, J.**, eds), pp. 87–130. Wallingford: CAB International.

Buck, A.H., Coakley, G., Simbari, F., McSorley, H.J., Quintana, J.F., Le Bihan, T., Kumar, S., Abreu-Goodger, C., Lear, M., Harcus, Y., Ceroni, A., Babayan, S.A., Blaxter, M., Ivens, A. and Maizels, R.M. (2014) Exosomes secreted by nematode parasites transfer small RNAs to mammalian cells and modulate innate immunity. *Nat. Commun.* **5**, 5488.

Chen, S., Lang, P., Chronis, D., Zhang, S., De Jong, W.S., Mitchum, M.G. and Wang, X. (2014) In planta processing and glycosylation of a nematode CLE effector and its interaction with a host CLV2-like receptor to promote parasitism. *Plant Physiol.* **167**, 262–272.

Cotton, J.A., Lilley, C.J., Jones, L.M., Kikuchi, T., Reid, A.J., Thorpe, P., Tsai, I.J., Beasley, H., Blok, V., Cock, P.J., Eves-van den Akker, S., Holroyd, N., Hunt, M., Mantelin, S., Naghra, H., Pain, A., Palomares-Rius, J.E., Zarowiecki, M., Berriman, M., Jones, J.T. and Urwin, P.E. (2014) The genome and life-stage specific transcriptomes of *Globodera pallida* elucidate key aspects of plant parasitism by a cyst nematode. *Genome Biol.* **15**(3), R43.

Danchin, E.G., Arguel, M.J., Campan-Fournier, A., Perfus-Barbeoch, L., Magliano, M., Rosso, M.N., Da Rocha, M., Da Silva, C., Nottet, N., Labadie, K., Guy, J., Artiguenave, F. and Abad, P. (2013) Identification of novel target genes for safer and more specific control of root-knot nematodes from a pan-genome mining. *PLoS Pathog.* **9**(10), e1003745.

De Boer, J.M., Yan, Y., Smant, G., Davis, E.L. and Baum, T.J. (1998) *In-situ* hybridization to messenger RNA in *Heterodera glycines*. *J. Nematol.* **30**, 309–312.

Fernandez, D., Tisserant, E., Talhinhas, P., Azinheira, H., Vieira, A., Petitot, A.S., Loureiro, A., Poulain, J., Da Silva, C., Silva, M.D. and Duplessis, S. (2012) 454-pyrosequencing of *Coffea arabica* leaves infected by the rust fungus *Hemileia vastatrix* reveals in planta-expressed pathogen-secreted proteins and plant functions in a late compatible plant–rust interaction. *Mol. Plant Pathol.* **13**, 17–37.

Gregory, T.R., Nicol, J.A., Tamm, H., Kullman, B., Kullman, K., Leitch, I.J., Murray, B.G., Kapraun, D.F., Greilhuber, J. and Bennett, M.D. (2007) Eukaryotic genome size databases. *Nucleic Acids Res.* **35**, D332–D338.

Gouzy, J., Carrere, S. and Schiex, T. (2009) FrameDP: sensitive peptide detection on noisy matured sequences. *Bioinformatics*, **25**, 670–671.

Haegeman, A., Mantelin, S., Jones, J.T. and Gheysen, G. (2012) Functional roles of effectors of plant-parasitic nematodes. *Gene*, **492**, 19–31.

Haegeman, A., Bauters, L., Kyndt, T., Rahman, M.M. and Gheysen, G. (2013) Identification of candidate effector genes in the transcriptome of the rice root knot nematode *Meloidogyne graminicola*. *Mol. Plant Pathol.* **14**, 379–390.

Hewitson, J.P., Grainger, J.R. and Maizels, R.M. (2009) Helminth immunoregulation: the role of parasite secreted proteins in modulating host immunity. *Mol. Biochem. Parasitol.* **167**, 1–11.

Huang, G., Gao, B., Maier, T., Allen, R., Davis, E.L., Baum, T.J. and Hussey, R.S. (2003) A profile of putative parasitism genes expressed in the esophageal gland cells of the root-knot nematode *Meloidogyne incognita*. *Mol. Plant–Microbe Interact.* **16**, 376–381.

Jaouannet, M., Perfus-Barbeoch, L., Deleury, E., Magliano, M., Engler, G., Vieira, P., Danchin, E.G., Da Rocha, M., Coquillard, P., Abad, P. and Rosso, M.N. (2012) A root-knot nematode secreted protein is injected into giant cells and targeted to the nuclei. *New Phytol.* **194**, 924–931.

Kawahara, Y., Oono, Y., Kanamori, H., Matsumoto, T., Itoh, T. and Minami, E. (2012) Simultaneous RNA-seq analysis of a mixed transcriptome of rice and blast fungus interaction. *PLoS One*, **7**(11), e49423.

Kawahara, Y., de la Bastide, M., Hamilton J.P., Kanamori, H., McCombie, W.R., Ouyang, S., Schwartz, D.C., Tanaka, T., Wu, J., Zhou, S., Childs, K.L., Davidson, R.M., Lin, H., Quesada-Ocampo, L., Vaillancourt, B., Sakai, H., Lee, S.S., Kim, J., Numa, H., Itoh, T., Buell, C.R. and Matsumoto, T. (2013) Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice*, **6**, 4.

Kyndt, T., Fernandez, D. and Gheysen, G. (2014) Plant-parasitic nematode infections in rice: molecular and cellular insights. *Annu. Rev. Phytopathol.* **52**, 135–153.

Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics*, **25**, 1754–1760.

Li, W. and Godzik, A. (2006) CD-HIT: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, **22**, 1658–1659

Lin, B., Zhuo, K., Wu, P., Cui, R., Zhang, L.H. and Liao, J. (2013) A novel effector protein, MJ-NULG1a, targeted to giant cell nuclei plays a role in *Meloidogyne javanica* parasitism. *Mol. Plant–Microbe Interact.* **26**, 55–66.

Lozano-Torres, J.L., Wilbers, R.H., Gawronski, P., Boshoven, J.C., Finkers-Tomczak, A., Cordewener, J.H., America, A.H., Overmars, H.A., Van't Klooster, J.W., Baranowski, L., Sobczak, M., Ilyas, M., van der Hoorn, R.A., Schots, A., de Wit, P.J., Bakker, J., Goverse, A. and Smant, G. (2012) Dual disease resistance mediated by the immune receptor Cf-2 in tomato requires a common virulence target of a fungus and a nematode. *Proc. Natl. Acad. Sci. USA*, **109**, 10 119–10 124.

Lozano-Torres, J.L., Wilbers, R.H., Warmerdam, S., Finkers-Tomczak, A., Diaz-Granados, A., van Schaik, C.C., Helder, J., Bakker, J., Goverse, A., Schots, A. and Smant, G. (2014) Apoplastic venom allergen-like proteins of cyst nematodes

modulate the activation of basal plant innate immunity by cell surface receptors. *PLoS Pathog.* **10**(12), e1004569.

Lu, S.W., Chen, S., Wang, J., Yu, H., Chronis, D., Mitchum, M.G. and Wang, X. (2009) Structural and functional diversity of CLAVATA3/ESR (CLE)-like genes from the potato cyst nematode *Globodera rostochiensis*. *Mol. Plant–Microbe Interact.* **22**, 1128–1142.

Marcilla, A., Trelis, M., Cortés, A., Sotillo, J., Cantalapiedra, F., Minguez, M.T., Valero, M.L., Sánchez del Pino, M.M., Muñoz-Antoli, C., Toledo, R. and Bernal, D. (2012) Extracellular vesicles from parasitic helminths contain specific excretory/secretory proteins and are internalized in intestinal host cells. *PLoS One*, **7**(9), e45974.

Matin, J.A. and Wang, Z. (2011) Next-generation transcriptome assembly. *Nat. Rev. Genet.* **12**, 671–682.

Mitchum, M.G., Wang, X., Wang, J. and Davis, E.L. (2012) Role of nematode peptides and other small molecules in plant parasitism. *Annu. Rev. Phytopathol.* **50**, 175–195.

Mitchum, M.G., Hussey, R.S., Baum, T.J., Wang, X., Elling, A.A., Wubben, M. and Davis, E.L. (2013) Nematode effector proteins: an emerging paradigm of parasitism. *New Phytol.* **199**, 879–894.

Miyawaki, K., Tabata, R. and Sawa, S. (2013) Evolutionarily conserved CLE peptide signaling in plant development, symbiosis, and parasitism. *Curr. Opin. Plant Biol.* **16**, 598–606.

Nakai, K. and Horton, P. (1999) PSORT: a program for detecting the sorting signals of proteins and predicting their subcellular localization. *Trends Biochem. Sci.* **24**, 34–35.

Nguyen, V.P., Bellafiore, S., Petitot, A.S., Haidar, R., Bak, A., Abed, A., Gantet P., Mezallira, I., de Almeida-Engler, J. and Fernandez, D. (2014) *Meloidogyne incognita*–rice (*Oryza sativa*) interaction: a new model system to study plant–root knot nematode interactions in monocotyledons. *Rice*, **7**, 23.

Opperman, C.H., Bird, D.M., Williamson, V.M., Rokhsar, D.S., Burke, M., Cohn, J., Cromer, J., Diener, S., Gajan, J., Graham, S., Houfek, T.D., Liu, Q., Mitros, T., Schaff, J., Schaffer, R., Scholl, E., Sosinski, B.R., Thomas, V.P. and Windham, E. (2008) Sequence and genetic map of *Meloidogyne hapla*: a compact nematode genome for plant parasitism. *Proc. Natl. Acad. Sci. USA*, **105**, 14 802–14 807.

Page, A.P., Stepek, G., Winter, A.D. and Pertab, D. (2014) Enzymology of the nematode cuticle: a potential drug target? *Int. J. Parasitol. Drugs Drug Resist.* **4**, 133–141.

Petersen, T.N., Brunak, S., von Heijne, G. and Nielsen, H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods*, **8**, 785–786.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J. and Glöckner, F.O. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596.

Quentin, M., Abad, P. and Favery, B. (2013) Plant parasitic nematode effectors target host defense and nuclear functions to establish feeding cells. *Front. Plant Sci.* **4**, 53.

Robinson, M.D., McCarthy, D.J. and Smyth, G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.

Rutter, W.B., Hewezi, T., Abubucker, S., Maier, T., Huang, G., Mitreva, M., Hussey, R. and Baum, T. (2014a) Mining novel effector proteins from the esophageal gland cells of *Meloidogyne incognita*. *Mol. Plant–Microbe Interact.* **27**, 965–974.

Rutter, W.B., Hewezi, T., Maier, T.R., Mitchum, M.G., Davis, E.L., Hussey, R.S. and Baum, T.J. (2014b) Members of the *Meloidogyne* avirulence protein family contain multiple plant ligand-like motifs. *Phytopathology*, **104**, 879–885.

Ruttkay-Nedecky, B., Nejdl, L., Gumulec, J., Zitka, O., Masarik, M., Eckschlager, T., Stiborova, M., Adam, V. and Kizek, R. (2013) The role of metallothionein in oxidative stress. *Int. J. Mol. Sci.* **14**, 6044–6066.

Schulz, M.H., Zerbino, D.R., Vingron, M. and Birney, E. (2012) Oases: robust *de novo* RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics*, **28**, 1086–1092.

Stergiopoulos, I. and de Wit, P.J.G.M. (2009) Fungal effector proteins. *Annu. Rev. Phytopathol.* **47**, 233–263.

Teixeira, P.J., Thomazella, D.P., Reis, O., do Prado, P.F., do Rio, M.C., Fiorin, G.L., José, J., Costa, G.G., Negri, V.A., Mondego, J.M., Mieczkowski, P. and Pereira, G.A. (2014) High-resolution transcript profiling of the atypical biotrophic interaction between *Theobroma cacao* and the fungal pathogen *Moniliophthora perniciosa*. *Plant Cell*, **26**, 4245–4269.

Thakur, K., Chawla, V., Bhatti, S., Swarnkar, M.K., Kaur, J., Shankar, R. and Jha, G. (2013) De novo transcriptome sequencing and analysis for *Venturia inaequalis*, the devastating apple scab pathogen. *PLoS One*, **8**(1), e53937.

Thorpe, P., Mantelin, S., Cock, P.J., Blok, V.C., Coke, M.C., Eves-van den Akker, S., Guzeeva, E., Lilley, C.J., Smant, G., Reid, A.J., Wright, K.M., Urwin, P.E. and Jones, J.T. (2014) Genomic characterisation of the effector complement of the potato cyst nematode *Globodera pallida*. *BMC Genomics*, 15, 923.

Trapnell, C., Pachter, L. and Salzberg, S.L. (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, **25**, 1105–1111.

Tremblay, A., Hosseini, P., Li, S., Alkharouf, N.W. and Matthews, B.F. (2013) Analysis of *Phakopsora pachyrhizi* transcript abundance in critical pathways at four time-points during infection of a susceptible soybean cultivar using deep sequencing. *BMC Genomics*, **14**, 614.

Wang, X., Allen, R., Ding, X., Goellner, M., Maier, T., de Boer, J.M., Baum, T.J., Hussey, R.S. and Davis, E.L. (2001) Signal peptide-selection of cDNA cloned directly from the esophageal gland cells of the soybean cyst nematode *Heterodera glycines*. *Mol. Plant–Microbe Interact.* **14**, 536–544.

Wasmuth, J.D. and Blaxter, M.L. (2004) prot4EST: translating expressed sequence tags from neglected genomes. *BMC Bioinformatics*, 5, 187.

Westermann, A.J., Gorski, S.A. and Vogel, J. (2012) Dual RNA-seq of pathogen and host. *Nat. Rev. Microbiol.* **10**, 618–630.

Yang, F., Li, W. and Jørgensen, H.J. (2013) Transcriptional reprogramming of wheat and the hemibiotrophic pathogen *Septoria tritici* during two phases of the compatible interaction. *PLoS One*, **8**(11), e81606.

Zeitoun-Ghandour, S., Leszczyszyn, O.I., Blindauer, C.A., Geier, F.M., Bundy, J.G. and Stürzenbaum, S.R. (2011) *Caenorhabditis elegans* metallothioneins: response to and defence against ROS toxicity. *Mol. Biosyst.* **7**, 2397–2406.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1 General annotation of *Meloidogyne graminicola* transcripts by Blast2GO.

Fig. S2 Mapped reads from the 14 RNA-sequencing (RNA-seq) libraries on the *de novo Meloidogyne graminicola* transcriptome. The library names (see Table 1 for description) are shown on the *x* axis. The percentages of mapped reads are shown on the *y* axis. The mapping percentage showed a maximum value for the pre-parasitic J2 samples, and varied from 2 to 16 days post-infection (dpi) following nematode development.

Fig. S3 Amino acid sequence of the candidate effector Mg4965. The putative signal peptide is underlined. Core CLE-like motifs are highlighted in yellow. The sequence logo shows the amino acid frequency of the CLE-like motifs.

Fig. S4 Expression profiles of *Meloidogyne graminicola* candidate effector genes determined by RPKM (reads per kilobase per million mapped reads) values and quantitative reverse transcription-polymerase chain reaction (RT-qPCR) analyses. Left, RPKM values obtained for various *M. graminicola* pre-parasitic and parasitic stages in rice plants. Right, relative fold change of gene expression levels in various *M. graminicola* parasitic stages in rice plants compared with the pre-parasitic J2 stage. *Mg-actin* was used as an internal control gene for normalization. Two independent biological assays were analysed (dark grey, replicate A; light grey, replicate B).

Table S1 General annotation of *Meloidogyne graminicola* predicted proteins by InterProScan. Distributions and frequencies of InterProScan and Pfam domains are presented for the complete proteome and for the predicted secretome.

Table S2 Putative secreted proteins from *Meloidogyne graminicola*. This table describes the sequence characteristics (size, cysteine content, nuclear prediction), annotations (BLAST results, InterProScan and Pfam domains) and RPKM (reads per kilobase per million mapped reads) values in the 14 RNA-sequencing (RNA-seq) libraries.

Table S3 Genes up- and down-regulated across the *Meloidogyne graminicola* life cycle in rice plants. Comparisons were performed between J2 and 2 days post-infection (dpi), 2 and 4 dpi, 4 and 8 dpi, 8 and 12 dpi, and 12 and 16 dpi. LogFC [log(fold change)], LogCPM [log(count per million)] and Padj (adjusted *P* value) were calculated by EdgeR. Genes encoding putative glycoside hydrolases, proteases or proteins involved in moulting or reproduction are presented in separate Excel sheets.

Table S4 Sequences of primers used for quantitative polymerase chain reaction (qPCR) and *in situ* hybridization experiments.